

Amendments to the Claims

This listing of claims will replace all prior versions, and listings, of claims in the application.

Listing of Claims:

1. (Cancelled)
2. (Cancelled)
3. (Cancelled)
4. (Cancelled)
5. (Cancelled)
6. (Cancelled)
7. (Cancelled)
8. (Cancelled)
9. (Cancelled)
10. (Cancelled)

11. (Cancelled)

12. (Cancelled)

13. (Cancelled)

14. (Cancelled)

15. (Cancelled)

16. (Cancelled)

17. (Cancelled)

18. (Cancelled)

19. (Cancelled)

20. (Cancelled)

21. (Cancelled)

22. (Cancelled)

23. (Cancelled)

24. (Cancelled)

25. (Cancelled)

26. (Cancelled)

27. (Cancelled)

28. (Cancelled)

29. (Cancelled)

30. (Cancelled)

31. (Cancelled)

32. (Cancelled)

33. (Cancelled)

34. (New) A method for supporting application workloads across multiple domains, the method comprising:

receiving a request from a client to execute a first application workload on a first server cluster at a first domain, the first server cluster at the first domain including a plurality of server nodes;

identifying a service level agreement negotiated with the client for the first application workload, the service level agreement specifying performance requirements for execution of the first application workload on the first server cluster at the first domain;

assigning a subset of the plurality of server nodes in the first server cluster at the first domain to execute the first application workload;

monitoring execution of the first application workload on the subset of server nodes assigned to execute the first application workload to determine whether the performance requirements for execution of the first application workload specified in the service level agreement are being met; and

responsive to a determination that the performance requirements for execution of the first application workload are not being met, sending a request to a second domain to assign one or more of a plurality of server nodes in a second server cluster at the second domain to the execution of the first application workload,

wherein a second application workload is executing on the second server cluster at the second domain while the first application workload is executing on the first server cluster at the first domain, and

wherein the request sent to the second domain specifies a number of server nodes requested, a duration in which the number of server nodes requested will be needed, and a dollar value associated with the request.

35. (New) The method of claim 34, further comprising:

monitoring execution of the first application workload on the subset of server nodes assigned to execute the first application workload to determine whether the performance requirements for execution of the first application workload specified in the service level agreement will continue to be met; and

responsive to a determination that the performance requirements for execution of the first application workload will not continue to be met, sending a request to the second domain to assign one or more of the plurality of server nodes in the second server cluster at the second domain to the execution of the first application workload.

36. (New) The method of claim 34, further comprising:

receiving a counter offer from the second domain, the counter offer specifying one or more of a different number of server nodes than the number of server nodes requested, a different duration for the number of server nodes requested, and a different dollar value for the request.

37. (New) The method of claim 36, further comprising:

sending a response to the second domain, the response indicating acceptance or rejection of the counter offer from the second domain.

38. (New) The method of claim 34, further comprising:

receiving a refusal from the second domain denying the request, the refusal from the second domain being based upon evaluation of an impact on a service level agreement negotiated for the second application workload executing on the second server cluster at the second domain if the request was granted.

39. (New) The method of claim 34, wherein responsive to a determination that the performance requirements for execution of the first application workload are not being met, the method further comprises:

sending a request to a third domain to assign one or more of a plurality of server nodes in a third server cluster at the third domain to the execution of the first application workload,

wherein a third application workload is executing on the third server cluster at the third domain while the first application workload is executing on the first server cluster at the first domain, and

wherein the request sent to the third domain specifies a number of server nodes requested, a duration in which the number of server nodes requested will be needed, and a dollar value associated with the request.

40. (New) The method of claim 39, further comprising:

receiving a first counter offer from the second domain, the first counter offer specifying one or more of a different number of server nodes than the number of server nodes requested, a different duration for the number of server nodes requested, and a different dollar value for the request; and

receiving a second counter offer from the third domain, the second counter offer specifying one or more of a different number of server nodes than the number of server nodes requested, a different duration for the number of server nodes requested, and a different dollar value for the request,

wherein what is specified in the first counter offer differs from what is specified the second counter.

41. (New) The method of claim 40, further comprising:

sending a first response to the second domain, the first response indicating acceptance or rejection of the first counter offer from the second domain; and

sending a second response to the third domain, the second response indicating acceptance or rejection of the second counter offer from the third domain.

42. (New) The method of claim 39, further comprising:

receiving a refusal from the third domain denying the request, the refusal from the third domain being based upon evaluation of an impact on a service level agreement negotiated for the third application workload executing on the third server cluster at the third domain if the request was granted.

43. (New) The method of claim 34, wherein the dollar value associated with the request is a penalty specified in the service level agreement negotiated for the first application workload.

44. (New) The method of claim 34, wherein the dollar value associated with the request is a payment amount for the number of server nodes requested.

45. (New) The method of claim 34, wherein the first application workload is a transaction application workload.

46. (New) The method of claim 45, wherein the transaction application workload comprises stock trades.

47. (New) The method of claim 34, wherein the second application workload is a parallel application workload.

48. (New) The method of claim 47, wherein the parallel application workload involves optimization of a stock portfolio.

49. (New) The method of claim 34, wherein the performance requirements for execution of the first application workload specified in the service level agreement comprises throughput requirements.

50. (New) The method of claim 34, wherein the performance requirements for execution of the first application workload specified in the service level agreement comprises response time requirements.

51. (New) The method of claim 34, wherein the performance requirements for execution of the first application workload specified in the service level agreement comprises availability requirements.

52. (New) The method of claim 34, wherein the performance requirements for execution of the first application workload specified in the service level agreement comprises downtime requirements.

53. (New) The method of claim 34, wherein the service level agreement negotiated with the client for the first application workload further specifies penalty functions when the service level agreement is violated.

54. (New) The method of claim 34, wherein monitoring execution of the first application workload on the subset of server nodes assigned to execute the first application workload comprises:

monitoring one or more of a transaction rate, a transaction response time, availability of a server node, and utilization of a server node.